Local stimulus disambiguation with global motion filters predicts adaptive surround modulation

Babette Dellen and Carme Torras

Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
Llorens i Artigas 4-6, 08028 Barcelona, Spain
{bdellen,torras}@iri.upc.edu

Abstract

Humans have no problem segmenting different motion stimuli despite the ambiguity of local motion signals. Adaptive surround modulation, i.e., the apparent switching between integrative and antagonistic modes, is assumed to play a crucial role in this process. However, so far motion processing models based on local integration have not been able to provide a unifying explanation for this phenomenon. This motivated us to investigate the problem of local stimulus disambiguation in an alternative and fundamentally distinct motion-processing model which uses global motion filters for velocity computation. Local information is reconstructed at the end of the processing stream through the constructive interference of global signals, i.e., inverse transformations. We show that in this model local stimulus disambiguation can be achieved by means of a novel filter embedded in this architecture. This gives rise to both integrative and antagonistic effects which are in agreement with those observed in psychophysical experiments with humans, providing a functional explanation for effects of motion repulsion.

Keywords: stimulus disambiguation, motion processing models, Fourier analysis, global filters, aperture problem, motion repulsion, constructive interference

1. Introduction

The classical receptive field (CRF) represents one of the most important concepts in neuroscience and has received large support since the work of Hubel and Wiesel (1968), who studied the response properties of neurons in the striate cortex of monkey to visual stimuli. The CRF has been defined as the region of the visual

Preprint submitted to Neural Networks *May 10, 2013*

space within which a stimulus evokes neuronal activity. Neurons in the primary visual cortex and area MT, which are considered to belong to the main motion processing pathway in most mammals, have been shown to have localized CRFs, and based on these findings, it was assumed that neural integration in the CRF is rather local than global, e.g. (Livingstone and Hubel, 1987).

However, the study of waves tells us that local information can also be represented in a distributed fashion by global signals. Fourier transformations use this principle to change from a space-time representation of the world to the frequency space, and vice versa. Without doubt, the principle of constructive interference[1] provides an equally plausible computational explanation for locality.

While both mechanisms can be used to predict local responses, using either one or the other necessarily results in two distinct, basic processing schemes, which are sketched in Fig. 1A-B. The two approaches incorporate global information differently. Hence, the study of locally ambiguous stimuli might shed some light on the important question of which processing architecture has been realized in the human visual system.

When characterizing the integrations used in classical models, the term *local* is used in the sense of *nearby*, meaning that integration is performed for nearby signals only, i.e., within a small area of the visual space. However, the term *local* is also used to describe a *specific location* in space. The superposition of global signals leads to the reconstruction of local image properties, despite the global nature of the preceding computations.

The aperture problem exemplifies how local integration limits the correct determination of stimulus velocity: When a moving line is seen through a small window such that the end points of the line are not visible, the motion of the line is ambiguous, because the observed temporal change in the window could have been caused by many different motions. Classical models attempt to disambiguate the stimulus using results obtained at neighboring locations (Fukushima, 1980; Huang et al., 2008a), assuming some locality-preserving interaction between local signals (see Fig. 1A). However, in case of multi-object stimuli, this strategy contains the risk that signals belonging to different entities are falsely combined, leading to a paradox: Before having disambiguated the stimuli, we cannot know

[1]The principle of superposition of waves states that when two or more waves are incident on the same point, the total amplitude at that point is equal to the sum the amplitudes of the individual waves. Constructive interference occurs when a crest of a wave meets a crest of another wave. Constructive interference is achieved in this work through inverse transformations, and both terms can be used synonymously here.

which local signals we can combine, and vice versa. This might explain why classical approaches so far have not been able to predict human perceptual responses to the stimuli shown in Fig. 4A on the basis of a single mechanism (Amano et al., 2009; Huang et al., 2008b).

In this work, the problem of stimulus disambiguation is investigated in the context of a framework based on constructive interference of global Fourier components, and a genuine solution to the problem is developed, made possible by a novel filter implemented in this architecture.

The paper is structured as follows. In Section 2, the underlying processing architecture based on constructive interference is introduced. Then, the results are presented, starting with a problem statement (Section 3), illustrating the problem of stimulus ambiguity using a typical example, and a solution is proposed (Section 4). Simulation results showing predictions of perceptual responses, including the motion-repulsion illusion are presented in Section 5. Finally, the results are discussed with respect to classical motion processing models (Adelson and Bergen, 1985; Fukushima, 1980; Huang et al., 2008a) based on local integration in Section 6.

2. Methods I: Mathematical model

We build upon a recent method and represent the visual stimulus through global Fourier components (Dellen and Wörgötter, 2011). This algorithm follows a processing structure based on the principle of constructive interference (see Fig. 1B), and thus provides an adequate ground for this case study.

Let $I(x, y, t)$ be a visual stimulus, where x and y define the spatial positions and t is the time. The spatial Fourier transform of $I(x, y, t)$ is then represented by

$$F(\mathbf{k}, t) = \iint_{-\infty}^{+\infty} \exp(i\mathbf{k}\mathbf{x}') I(\mathbf{x}', t) d\mathbf{x}' \quad , \tag{1}$$

where $\mathbf{k} = (k_x, k_y)$ is the spatial frequency vector, and $\mathbf{x}' = (x', y')$. Accordingly, the spatiotemporal Fourier transform is defined as

$$F(\mathbf{k}, \omega) = \int_{-\infty}^{+\infty} F(\mathbf{k}, t') \exp(i\omega t') dt' \quad , \tag{2}$$

where ω is a temporal frequency.

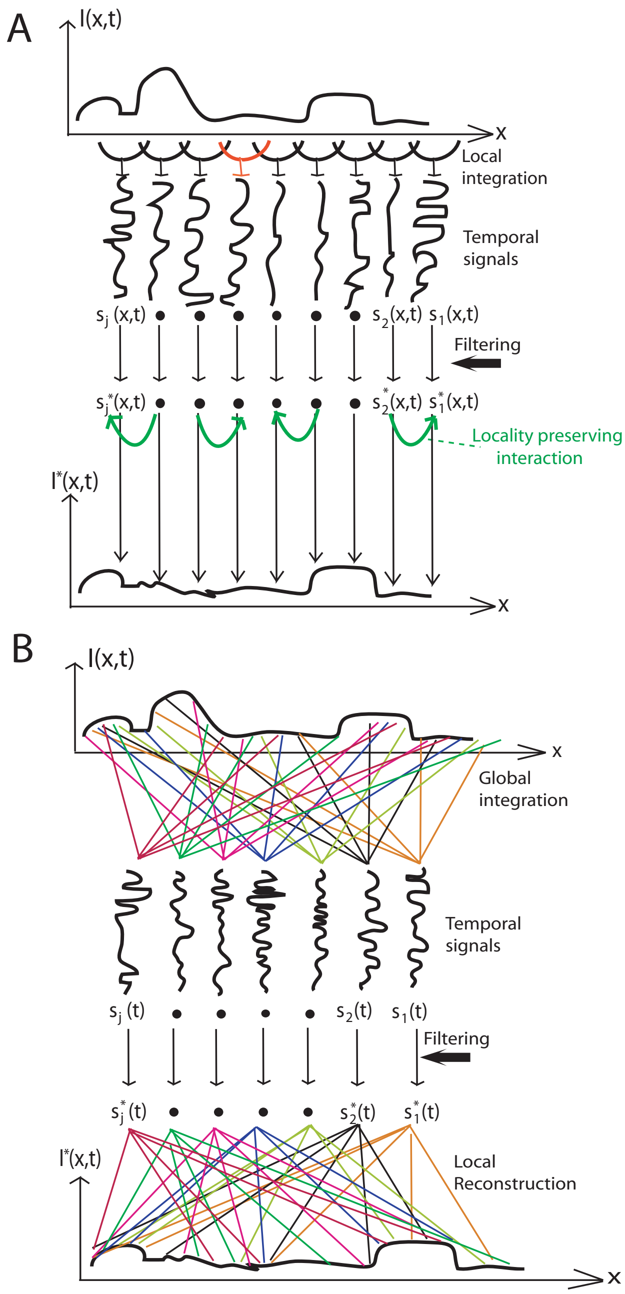

Figure 1: A. Classical models assume a processing architecture using local integration inside small patches of the stimulus $I(x,t)$ to compute local stimulus properties. The resulting signals $s_i(x,t)$ depend thus on the location of the respective patch and encode relevant local stimulus features, as for example motion. Global information is incorporated by enabling a locality-preserving interaction between local signals. The final responses $s_i^*(x,t)$ then map the input space in a trivial way. B. The principle of constructive interference defines an alternative computational strategy. Global integrations generate structured, temporal signals $s_i(t)$, which, taken individually, do not provide information about location. When superimposing the signals in an appropriate manner, spatial information can be reconstructed fully. Specific stimulus features can be extracted through temporal filtering of the global signals before reconstruction.

To compute image motion, we can filter the Fourier components temporally with

$$b(\mathbf{k}, \omega, \mathbf{v}_p, t) = a(\mathbf{k}, \omega, \mathbf{v}_p) \exp{(i\omega t)} \quad , \tag{3}$$

where

$$a(\mathbf{k}, \omega, \mathbf{v}_p) = \exp{[-(\omega - \mathbf{k}\mathbf{v}_p)^2/(|\mathbf{k}|^2\sigma^2)]}. \tag{4}$$

Here, $\mathbf{v}_p = (v_p \cos\Theta, v_p \sin\Theta)$ is the preferred (or search) velocity (Dellen and Wörgötter, 2011), defined through a preferred motion direction Θ and a preferred speed v_p. The parameter σ determines how strongly the motion constraint equation is enforced. We use $\sigma = 0.2$ pixels/frame. With a larger σ, a wider range of values is sampled.

Applying the inverse Fourier transform to the filtered Fourier space allows us to return in real space and reconstruct those parts of the image that move with a velocity $\mathbf{v}_p$, yielding a local response

$$r_L(\mathbf{x}, t, v_p, \Theta) = \iiiint_{-\infty}^{+\infty} \alpha \, b(\mathbf{k}, \omega, \mathbf{v}_p, t - t')F(\mathbf{k}, t')d\mathbf{k}d\omega dt' \tag{5}$$

$$= \iiint_{-\infty}^{+\infty} \alpha \, \beta \, a(\mathbf{k}, \omega, \mathbf{v}_p)F(\mathbf{k}, \omega)d\mathbf{k}d\omega \tag{6}$$

with the inverse transformation factors $\alpha = \exp{(-i\mathbf{k}\mathbf{x})}$ and $\beta = \exp{(-i\omega t)}$. Using the velocity filter, a global response can be defined in an analog manner as

$$r_G(v_p, \Theta) = \iiint_{-\infty}^{+\infty} |a(\mathbf{k}, \omega, \mathbf{v}_p)F(\mathbf{k}, \omega)|d\mathbf{k}d\omega. \tag{7}$$

The term *global* requires some further clarification. In this study, we integrate over the entire stimulus to obtain the global response. For larger images, both local and global responses could be alternatively computed by integrating only over a part of the stimulus, for example by using Gabor filters (Heeger, 1988).

To illustrate the properties of the filter presented in Eq. 3, we computed the local responses for a synthetic stimulus consisting of two squares moving with

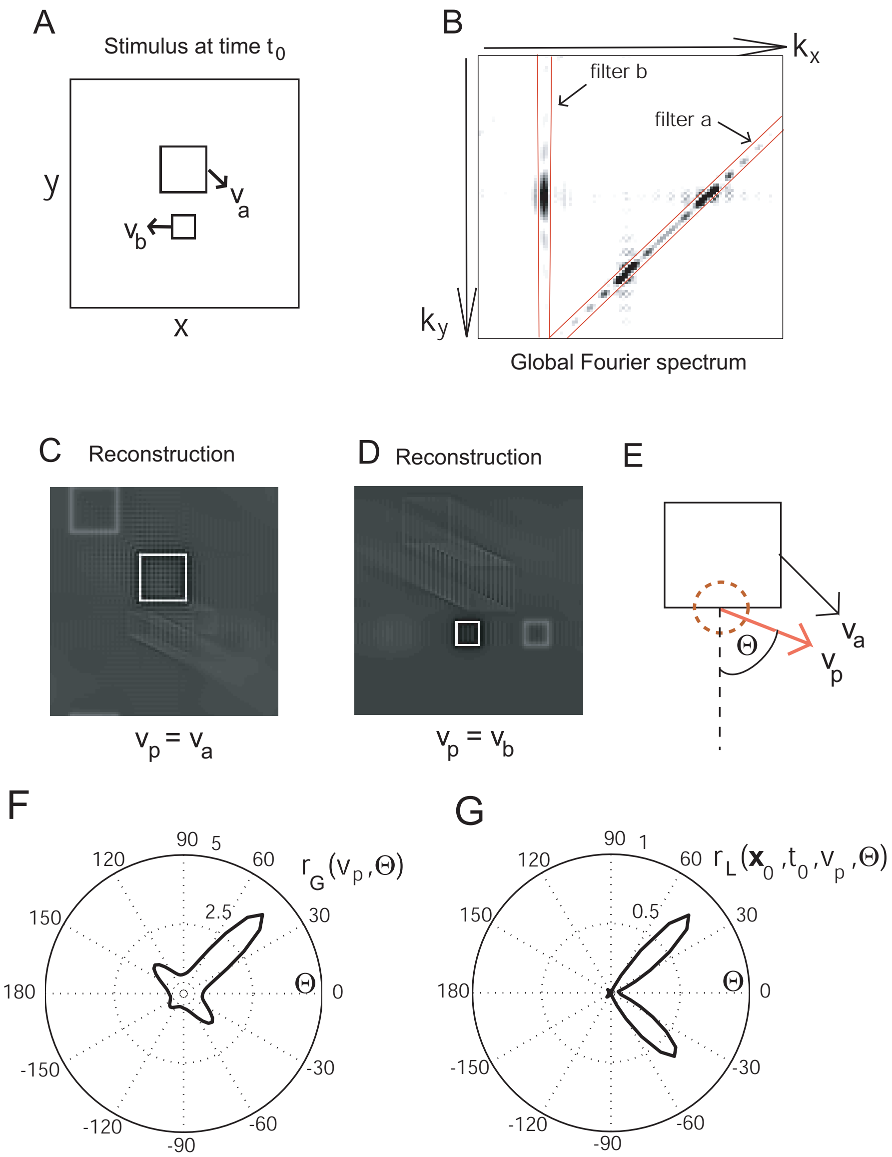

Figure 2: A. Schematic of the stimulus consisting of two moving squares at time t_0. B. Slice through the global Fourier spectrum. Boundaries of the velocity filters with $\mathbf{v}_p = \mathbf{v}_a$ and $\mathbf{v}_p = \mathbf{v}_b$ are schematically represented in red. The origin of the spectrum has been shifted to the center. C. Reconstruction $r_L(\mathbf{x}, t, v_p, \Theta)$ with $\mathbf{v}_p = \mathbf{v}_a$. D. Reconstruction $r_L(\mathbf{x}, t, v_p, \Theta)$ with $\mathbf{v}_p = \mathbf{v}_b$. E. Object velocity, preferred velocity and the respective preferred angle Θ. F-G. Global and local responses as a function of Θ as polar plots, respectively.

different velocities $\mathbf{v}_a = (1, 1)$ pixels/frame and $\mathbf{v}_b = (-1, 0)$ pixels/frame (see Fig. 2A) [2]. A slice through the global Fourier spectrum is shown in Fig. 2B together with the velocity filters with $\mathbf{v}_p = \mathbf{v}_a$ and $\mathbf{v}_p = \mathbf{v}_b$. The respective local responses obtained with the filters $r_L(\mathbf{x}, t, v_p, \Theta)$ are shown in Fig. 2C-D, respectively, demonstrating that the local properties of the object moving in the preferred direction can be reconstructed with high precision, while objects moving in the non-preferred direction are eliminated. Local velocity can be estimated from the local responses obtained for different preferred velocities.

In this model, a non-causal temporal filter is used. Future space-time events could be truncated by using a causal filter. As long as the same number of frames (total integration time) is used in the computations, this would lead to the same results, at least for the stimuli considered in this paper (constant motions). However, when shortening the length of the motion sequences, responses will naturally become more ambiguous, rendering the use of additional mechanisms for disambiguation (as described in Section 4) even more important.

3. Results I: Analyzing ambiguous stimuli

In the following, we analyze the problem of local stimulus disambiguation in the mathematical models presented in the previous section.

Let us study the case of a square of size 20×20 pixels moving diagonally downwards in more detail (see Fig. 2E). In Fourier space, the stimulus is represented by a line oriented perpendicular to the stimulus velocity $\mathbf{v}_a$, corresponding to the motion-constraint line of the stimulus (as shown in Fig. 2B). A velocity filter with $\mathbf{v}_p = \mathbf{v}_a$ extracts all Fourier components which are close to this line (see Fig. 2B, red lines for filter a), allowing full reconstruction of the object.

Let us now compute the global response for different preferred motion directions Θ for a constant preferred speed $v_p = \sqrt{2}$ pixels/frame, as illustrated in Fig. 2E. As shown in Fig. 2F, the largest global response is obtained for $\Theta = 45$ deg, which is the direction of motion of the square. However, the global response curve contains two subsidiary peaks at 45 ± 90 degs. This is because the spatial frequency components at the intersection with the k_x and k_y axes of the constraint line are attenuated due to the specific structure of the object. Responses

[2] Computer simulations have been performed using Matlab's built-in functions fftn and fft2 using default conditions. Image sequences representing stimuli consisted of 20 frames of size 100×100 pixels. The parameter Θ was sampled in steps of 5 degs.

at 45 ± 45 degs are diminished because the component speed at these angles is smaller than the preferred speed of the used motion filter.

We sum the local responses $r_L(\mathbf{x}, t, v_p, \Theta)$ inside a small area with radius $r = 3$ pixels around the center of the lower contour of the square at $\mathbf{x}_0 = (50, 50)$ pixels and time t_0 (Fig. 2E, dotted red circle) for different motion directions Θ to get an average of the response characteristics in the center. We obtain a strong response for the true direction of motion with $\Theta = 45$ deg. However, as stated before, the spatial frequency components with $k_y = 0$ are contributing proportionally more to the reconstruction of the point than others, leading to a pronounced subsidiary peak at $\Theta = -45$ deg (see Fig. 2G). Nevertheless, the peak at $\Theta = 45$ deg is larger than the one at $\Theta = -45$ deg, demonstrating that the ambiguity problem is partly solved by the use of the principle of constructive interference. This is because in the model also low spatial-frequency components, carrying valuable information about the true direction of motion of long contours, contribute to the reconstruction of local information when computing local responses according to Eq. 6.

We could now compute the correct velocity of the contour point by selecting the angle for which the largest response is obtained, as it was done in Dellen and Wörgötter (2011). However, such a solution would be rather fragile because the difference between the two peaks is small. Noise or other factors would make this approach fail easily. Noise can enter the system through various channels, e.g., image noise, failure of the underlying motion assumptions, and numerical issues (Simoncelli et al., 1991).

By looking at the global response, we observe that the local response could be disambiguated more robustly by multiplying it with its global counterpart, yielding

$$r(\mathbf{x}, t, v_p, \Theta) = r_L(\mathbf{x}, t, v_p, \Theta) r_G(v_p, \Theta). \tag{8}$$

However, this operation would affect filter responses to other objects moving in a direction close to the global direction of motion, because the respective filters are bound to overlap in global Fourier space.

To illustrate this, we computed the product of the filters that are selective for different directions of motion as a function of their relative angle (see Fig. 3A, black lines). For opposing motions, the filter responses can be considered as independent, but for small angles, the overlap is severe and cannot be neglected.

The most direct way to resolve this problem would be to transform the incoming stimuli such that the relative angle between their motion is increased.

Once this is achieved, the operation defined by Eq. 8 can be safely applied. This may explain motion-repulsion effects that have been observed in psychophysical experiments (Marshak and Sekuler, 1979; Braddick et al., 2002). The overlap between filters in global Fourier space should thus be predictive for the strength of the repulsion effect, and, indeed, a similar dependency on the relative angles has been observed in psychophysical experiments by Marshak and Sekuler (1979) (red diamonds for subject w.m. and green circles for the average) and Braddick et al. (2002) (blue squares), as illustrated in Fig. 3A, for angles larger than $10-20$ degrees. Below this threshold, stimuli are considered identical, and no disambiguation, hence repulsion, is required.

For highly ambiguous stimuli, e.g., bar-with-square (see Fig. 4A, middle panel), responses are expected to be even more affected by the overlap between motion filters because their Fourier energies are concentrated right at the intersection point of the filters. By measuring the overlap between filtered stimulus components, we can provide a measure of the expected ambiguity of the response. Let us consider a stimulus consisting of a long bar moving straight downwards and an open square moving at an angle of 45 degrees relative to the bar (as illustrated in Fig. 4A, middle panel). With decreasing bar length, this overlap becomes smaller, as shown in Fig. 3B (black circles). Here, the normalized overlap is plotted as a function of the normalized bar length. Huang et al. (2008b) measured the strength of repulsion for two different bar lengths for the same type of stimulus (red squares). The overlap of the filtered stimuli is predictive for the strength of the repulsion effects. Huang et al. (2008b) also measured the strength of the repulsion effect for a dots-with-square stimulus (as illustrated in Fig. 4A, right panel). For this stimulus, we find a small overlap between the dots and the square (dotted black line) compared to the long bar, which also agrees with experimental measurements (red dotted line) (Huang et al., 2008b).

The overlap of motion filters affects the responses computed in the model. This is demonstrated for the stimuli depicted in Fig. 4A. The square, representing the surround stimulus, is moving in all three cases with a global velocity $\mathbf{v}_G = (1, 1)$ pixels/frame. In the bar-with-square and the dots-with-square stimulus, the lower contour of the square was replaced with a bar and a dot pattern, respectively, moving straight downwards with a velocity of $(0, 1)$ pixels/frame, thus making an angle of 45 deg with the global direction of motion. The responses $r_L(\mathbf{x}_0, t_0, v_p, \Theta)$ in the area of the center stimuli (inside the red dotted circle) are affected by the multiplication with the global signals, as shown in Fig. 4B-C. Multiplication with the global responses correctly disambiguates the contour of the translating square, but $r(\mathbf{x}_0, t_0, v_p, \Theta)$ wrongly suggests a local motion with

$\mathbf{v}_G$ in case of the bar-with-square stimulus. This is caused by the overlap of the filters positioned at 45 and 0 degs. The dots-with-square stimulus is less affected than the bar-with-square stimulus, because the respective initial local response is not ambiguous (see Fig. 4B, right panel).

4. Methods II: Proposed mathematical solution

As explained above, the overlap between motion stimuli can be reduced by applying a transformation that increases the angular distances between nearby motion stimuli. This transformation should leave the signals belonging to the surround stimulus unaffected, while shifting the motion direction of the center stimulus away from the global direction. This can be achieved by translating the stimulus $I(\mathbf{x},t)$ for all times t by a distance $-\mathbf{v}_G(t-t_0)/2$ with respect to a given time point t_0. This corresponds to a phase shift in Fourier space, yielding

$$F_s^*(\mathbf{k},\omega) \quad = \quad \int_{-\infty}^{+\infty} F(\mathbf{k},t)\exp\left[i\mathbf{k}\mathbf{v}_G(t-t_0)/2\right]\exp\left(i\omega t\right)dt \quad , \tag{9}$$

which can be understood as a straightforward extension of the temporal filter defined in Eq. 3 by a factor $\exp\left[i\mathbf{k}\mathbf{v}_G(t-t_0)/2\right]$. Here, $\mathbf{v}_G$ is assumed to be the global velocity of the stimulus, which can be determined by finding the preferred velocity for which the global response has its maximum. The velocity-shift vector $\mathbf{v}_G/2$ was chosen to obtain similar repulsion effects as observed in psychophysical experiments.

The change in absolute speed caused by the transformation can be corrected by doubling the temporal frequency according to

$$F_s(\mathbf{k},t) = \int_{-\infty}^{+\infty} F_s^*(\mathbf{k},\omega)\exp\left(-i2\omega t\right)d\omega. \tag{10}$$

Alternatively, the preferred velocity of the filter given by Eq. 3 could be bisected. Replacing $F(\mathbf{k},t')$ by $F_s(\mathbf{k},t')$ in Eq. 5 (and accordingly in Eqs. 6 and 7) defines the (shifted) response

$$r_s(\mathbf{x},t,v_p,\Theta) = r_{L,s}(\mathbf{x},t,v_p,\Theta)r_{G,s}(v_p,\Theta). \tag{11}$$

for $t=t_0$ with $t_0=10$ frames.

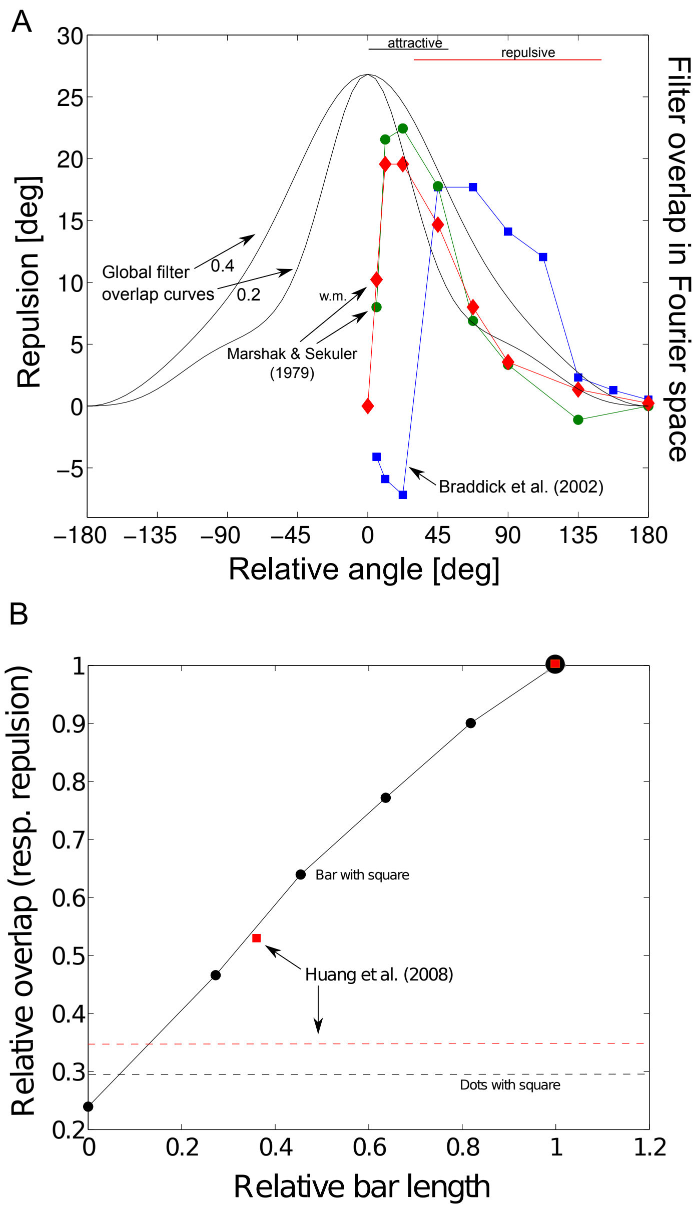

Figure 3: A. The normalized and scaled overlap of motion filters in Fourier space is shown as a function of the relative angle between motion directions for two different values of σ (solid black lines) and $v_p = 0.5$ pixels/frame. For small angles the overlap is severe, and the most effective way to resolve the problem is to transform the incoming stimuli such that the relative angle between their motion directions is increased, predicting repulsion effects observed in psychophysical experiments (Marshak and Sekuler, 1979; Braddick et al., 2002). For angles smaller than $10 - 20$ deg, the stimuli can be considered identical, and no disambiguation is required. Consequently, no repulsion is observed. B. The normalized overlap between a bar and square (see Fig. 4A, middle panel) for different bar sizes (normalized) is shown as a function of the normalized length of the bar (black circles). The overlap (reflecting the ambiguity of the stimulus) increases with increasing bar length, thus more repulsion would be needed to disambiguate the stimulus, which agrees with psychophysical data Huang et al. (2008b) (red squares). Note that the red square and the black circle at 1 are plotted at the same position.

5. Results II: Predicting perceptual responses in the computer simulation

We will show that by changing to a reference system that is moving with $-v_G/2$, the motion directions of the objects can be sufficiently disentangled for the stimuli shown in Fig. 4A.

The transformation described in Eqs. 9-11 leaves the contour of the square unaffected, and we obtain the same response in the non-shifted case as in the shifted case (see Fig. 4C-D, respectively). For the bar-with-square stimulus we obtain a quite distinct response for $r_s(\mathbf{x}, t_0, v_p, \Theta)$ compared to $r(\mathbf{x}, t_0, v_p, \Theta)$. For example, by subtracting r from r_s, a distinguished peak at -45 deg can be identified. Similar results are obtained for the dots-with-square stimulus (Fig. 4C-D, right panel). Based on the responses in both the non-shifted and the shifted case, we can now discriminate between the different input stimuli.

Different ways for estimating local motion directions can be devised which allow differentiating between motion-defined objects. For example, after subtracting r partially from r_s, the resulting peak can be easily identified by finding the maximum response, yielding

$$\Theta_{\text{est}}(\mathbf{x}, t_0) \quad = \quad \arg\{\max[r_s(\mathbf{x}, t_0, v_p, \Theta) - \epsilon r(\mathbf{x}, t_0, v_p, \Theta)]\} \quad , \tag{12}$$

where ϵ should be chosen such that the shifted responses do not get diminished entirely. The predicted motion direction fields computed for $\epsilon = 0.3$ are shown in the left panels of Fig. 5A-C for the three different stimuli. The corresponding predicted directions for the center stimulus are shown as blue arrows in the right panels. The predicted direction depends on the amount of shift introduced in Eq.9. However, to correctly disambiguate the dots-with-square stimulus, a shift smaller than $-\mathbf{v}_G/2$ would have been sufficient, as can be concluded from the results shown in Fig. 4B-C (right panel) and Fig. 3B. Repulsion has been observed for the analyzed stimuli in psychophysical experiments (Amano et al., 2009; Huang et al., 2007, 2008b), and similar effects have been observed in electrophysiological experiments (Allman et al., 1985; Xiao et al., 1995; Tanaka et al., 1986). In an advanced system, the shifting operation should be optimized in way that minimizes the required amount of shift while still succeeding in disambiguating the stimulus.

The observed shift of the center stimulus away from the global motion direction of the surround is known as the motion-repulsion illusion (Marshak and Sekuler, 1979). Visual illusions provide valuable insights into the functional

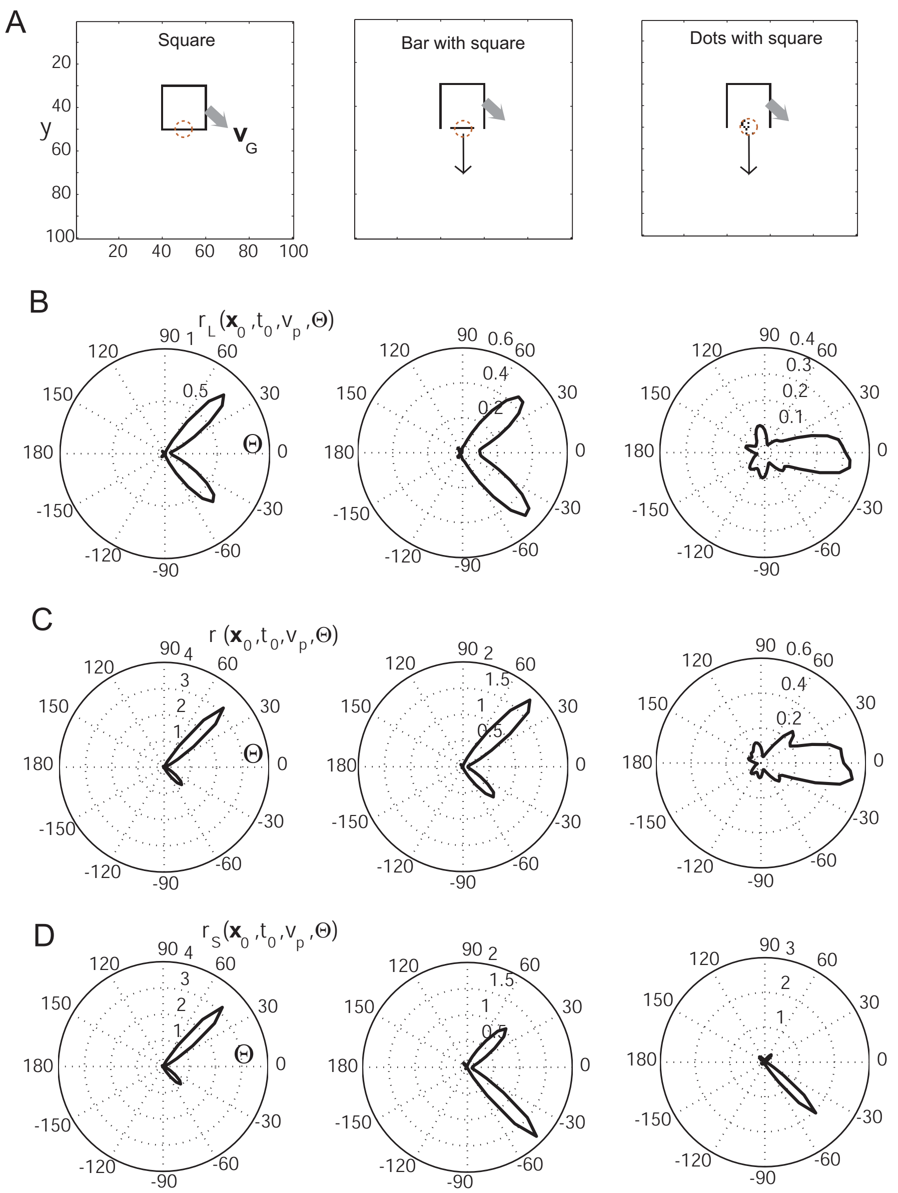

Figure 4: Stimulus disambiguation in the computer simulation. A. Schematics of the visual stimuli. B. Local responses $r_L(\mathbf{x}_0, t_0, v_p, \Theta)$ as a function of Θ. C. Local responses multiplied by the global response, giving $r(\mathbf{x}_0, t_0, v_p, \Theta)$. D. Local responses multiplied by the global response using $F_s(\mathbf{k}, t)$, giving $r_s(\mathbf{x}_0, t_0, v_p, \Theta)$. Using both responses $r(\mathbf{x}_0, t_0, v_p, \Theta)$ and $r_s(\mathbf{x}_0, t_0, v_p, \Theta)$, the square stimulus can be clearly distinguished from the bar-with-square and the dots-with-square stimulus. Videos of the stimuli can be downloaded from *http://www.iri.upc.edu/people/bdellen/Illusions.html.*

mechanism in our brain, and their correct prediction is an important cue when developing models of visual function (Gregory, 1966; Martineau et al., 2009).

In this study, we restricted our analysis to the most relevant case, i.e., $v_p = v_G$, for which the ambiguity problem and its solution is demonstrated best. Due to the multiplication by the global response, responses at preferred speeds that do not match any of the global motions of the stimulus are small, and thus can be neglected in the analysis. This is best exemplified for the square stimulus. For $v_p = 1$ pixels/frame we will obtain a peak at $\Theta = 0$ deg in the local response for a point on the lower contour. However, this peak will be largely diminished when multiplying it by the global response. The same is true for the shifted case. For the dots-with-square stimulus, the peak at $v_p = 1$ pixels/frame will be diminished neither for the non-shifted nor the shifted responses because a global motion with $v_G = 1$ pixels/frame is present, thus being in agreement with the results already obtained for $v_p = v_G$.

6. Discussion

In this work, we presented a solution to the problem of local stimulus disambiguation embedded in a motion-processing structure based on the principle of constructive interference. By applying a novel filter to the global Fourier transform of the stimulus, resembling a "stretching of the Fourier space", the overlap of the Fourier components belonging to the different objects is decreased, and long contours can be successfully disambiguated by multiplying the local responses of motion filters with their global counterpart. Motion-direction estimates computed from the motion-filtered responses correctly predict human perceptual responses to stimuli previously used in experiment, displaying both integrative and antagonistic effects (adaptive modulation), which so far have been assumed to originate from separate mechanisms.

Could an equivalent shifting mechanism as proposed in this paper be used to resolve local ambiguities in a model based on local integration? Let us look at the bar-with-square stimulus again. If we translate the stimulus with $-\mathbf{v}_G/2$, the bar would be moving diagonally downwards and to the left. Hence, the local response would still be affected by the aperture problem and be locally indistinguishable from the contour motion of the translating square. Our answer must thus be in the negative. Perceptual measurements clearly show that our visual system successfully disambiguates the motion of the bar, providing strong support for a motion processing structure based on constructive interference at the expense of local-integration models. In the proposed framework, the different stimuli can be

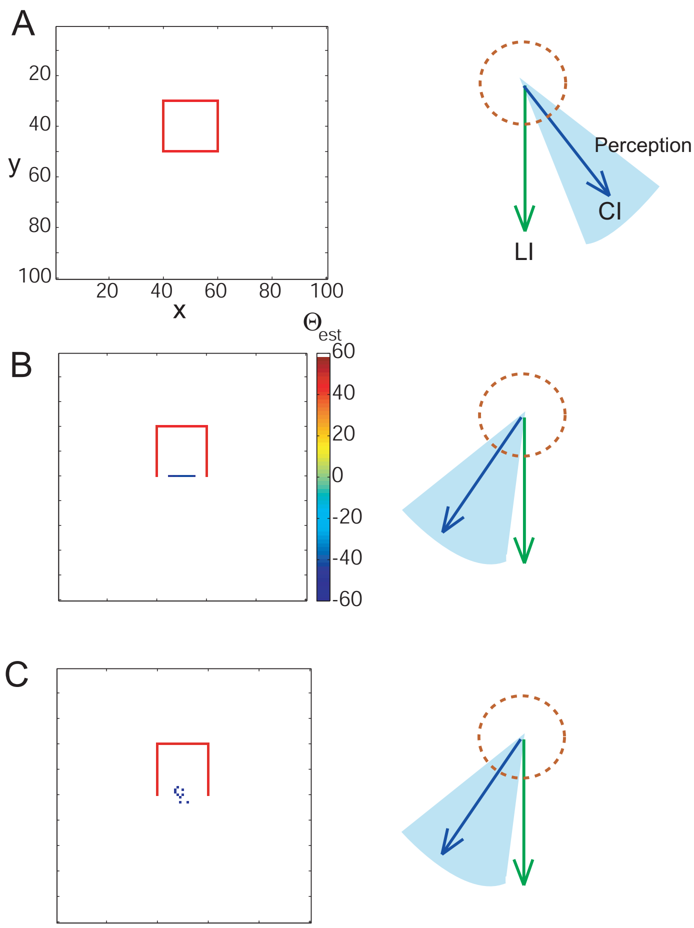

Figure 5: Left panels: Estimated local motion directions for the stimuli shown in Fig. 4. Right panels: The prediction from our model using constructive interference (CI) (based on a constant shift of $-\mathbf{v}_G/2$) is shown as a blue vector. The prediction from local integration (LI) (green arrow) is shown for comparison. The perceptually observed response range is illustrated using shaded light blue areas (Huang et al., 2007, 2008b). The strength of repulsion varies with subjects and stimuli, and inter-subject variability would be too detailed to be displayed here. Integrative modulation is observed for the square stimulus (A), while for the bar-with-square and the dots-with square stimuli antagonistic effects are observed (B-C).

disambiguated because low spatial frequencies can enter the computations without impairing the locality of the solution, which constitutes a major difference compared to local-integration models.

The proposed framework may further explain discrepancies between measured perceptual responses and neuronal responses in area MT that have been observed for the bar-with-square stimulus (Huang et al., 2008b). While perceptual responses for the bar are clearly antagonistic, the response of MT neurons shifted on average slightly towards the motion of the surround. When computing the weighted average of the shifted responses (see Fig. 4D, middle panel), we observe a similar effect, whose strength may vary depending on the shift employed. Neural responses in area MT could represent this intermediate stage in our processing stream, while the final perceptual response is derived from the whole population.

Based on our analysis of the stimulus-disambiguation problem, we further predict that the shifting mechanism explored in this paper needs only to be employed for small angles with the global motion direction, since the overlap of the global filters is falling with increasing angles. This decrease matches the perceptually measured decay of the motion-repulsion effect for large angles (see Fig. 3) (Marshak and Sekuler, 1979; Braddick et al., 2002). Direction repulsion has also been observed for transparent stimuli (Braddick et al., 2002; Benton and Curran, 2003; Treue et al., 2000), and it has been suggested that it originates from global mechanisms rather than local ones (Benton and Curran, 2003), which is in agreement with our work. In global Fourier space, transparent stimuli share a lot of properties with non-transparent stimuli, and could be treated using the proposed framework by allowing more than one motion to be computed for each point.

Fourier transforms have been used for motion analysis before (Watson and Ahumada, 1983; Porat and Friedlander, 1990; Huang and Chen, 1995; Dellen and Wörgötter, 2011). The algorithm of Porat and Friedlander (1990) uses directional filtering in the frequency domain to find dim targets. Velocity filtering in the Fourier domain and inverse transformations have also been used to compute image velocities (Dellen and Wörgötter, 2011). Both algorithms utilize inverse transformations to reconstruct position information, and thus follow a processing architecture based on constructive interference. Most other algorithms operating in the space-time frequency domain, however, derive spatio-temporal position by confining the analysis to a small area of the visual space (Adelson and Bergen, 1985; Huang and Chen, 1995; Fleet and Jepson, 1990; Bayerl and Neumann, 2004). Phase-based techniques apply Gabor filters to the image and use the phase of the Gabor filter response to estimate local velocity, having the advantage that the phase of Gabor filters is more robust to noise than, e.g., image gradients (Fleet and Jepson, 1990; Felsberg, 2004).

Hybrid models between the different processing architectures presented in Figs. 1A and 1B are also thinkable. For example, Gabor functions could be used instead of using Fourier transformations (Daugman, 1988). In this case, position would have to be derived through inverse Gabor transformations. This would allow eliminating information from distracting objects that are far apart.

Our work indicates that the principle of constructive interference might be of fundamental importance for understanding visual processing in the brain. By using constructive interference, the computation of local values can be postponed until the end of the processing stream is reached. This satisfies the holistic paradigm (Wertheimer, 1912; Koffka, 1935) in the sense that the solution is found by considering the entire input simultaneously, and *not* by combining elementary solutions computed on subsets of the data. Moreover, no distinction has to be made anymore between center and surround to explain the observed phenomenon, providing a great simplification of the receptive-field concept. Filters embedded in this architecture could be adapted through learning mechanisms without changing the principal structure of the system.

We finally stress that the presented work exclusively investigates the implications of computational principles of visual processing, and not explicit neuronal models implementing such principles.

7. Acknowledgements

B.D. acknowledges support from the Spanish Ministry for Science and Innovation through a Ramon y Cajal program.

References

Adelson, E., Bergen, J., 1985. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A (2), 284–298.

Allman, J., Miezin, F., McGuinness, E., 1985. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu. Rev. Neurosci. 8, 407–430.

Amano, K., Edwards, M., Badcock, D., Nishida, S., 2009. Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-element stimulus. J. of Vision 9, 1–25.

Bayerl, P., Neumann, H., 2004. Disambiguating visual motion through contextual feedback modulation. Neural Comput. 16 (10), 2041–2066.

Benton, C., Curran, W., 2003. Direction repulsion goes global. Current Biology 13, 767–771.

Braddick, O., Wishart, K., Curran, W., 2002. Directional performance in motion transparency. Vision Research 42, 1237–1248.

Daugman, J., 1988. Complete 2-d gabor transforms by neural networks for image analysis and compression. IEEE Trans. on Acoustics, Speech, and Signal Proc. 36 (7), 1169 – 1179.

Dellen, B., Wörgötter, F., 2011. A local algorithm for the computation of image velocity via constructive interference of global fourier components. Int. J. Comp. Vis. 92, 53–70.

Felsberg, M., 2004. Optical flow estimation from monogenic phase. IWCM, 1–13.

Fleet, D., Jepson, A., 1990. Computation of component image velocity from local phase information. Int. J. Comp. Vis. 5 (1), 77–104.

Fukushima, K., 1980. Neocognitron: a self organizing neural networkd for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36 (4), 193–202.

Gregory, R., 1966. Eye and brain: The psychology of seeing. London: Weidenfeld and Nicolson.

Heeger, D., 1988. Optical flow using spatiotemporal filters. Int. J. Comp. Vis. 1 (4), 279–302.

Huang, C.-L., Chen, Y.-T., 1995. Motion estimation method using a 3d steerable filter. Image and Vis. Comp. 13 (1), 21–32.

Huang, W., Jiao, L., Jia., J., 2008a. Modeling contextual modulation in the primary visual cortex. Neural Networks 21 (8), 1182–1196.

Huang, X., Albright, T., Stoner, G., 2007. Adaptive surround modulation in cortical area Mt. Neuron 53, 761–770.

Huang, X., Albright, T., Stoner, G., 2008b. Stimulus dependency and mechanisms of surround modulation in cortical area Mt. J. Neurosci. 8, 13889–13906.

Hubel, D. H., Wiesel, T. N., 1968. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195, 215–243.

Koffka, K., 1935. Principles of Gestalt Psychology. New York: Harcourt Brace.

Livingstone, M., Hubel, D., 1987. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J. Neurosci. 7, 3416–3468.

Marshak, W., Sekuler, R., 1979. Mutual repulsion between moving visual targets. Science 205, 1399–1401.

Martineau, P., Aguilar, M., Glass, L., 2009. Predicting perception of the wagon wheel illusion. Phys. Rev. Lett. 103, 028701.

Porat, B., Friedlander, B., 1990. A frequency domain algorithm for muliframe detection and estimation of dim targets. IEEE Trans. Pattern Anal. Mach. Intell. 12 (4), 398–401.

Simoncelli, E., Adelson, E., Heeger, D., 1991. Probability distributions of optical flow. IEEE Intl. Conf. Comp. Vis. Patt. Recog., 310–315.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., Iwai, E., 1986. Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. J. Neurosci. 1986, 134–144.

Treue, S., Hol, K., Rauber, H., 2000. Seeing multiple directions of motion-physiology and psychophysics. Nat Neurosci. 3 (3), 270–276.

Watson, A., Ahumada, A., 1983. A look at motion in the frequency domain. New York: Association for Computing Machinery.

Wertheimer, G., 1912. Experimentelle Studien über das Sehen von Bewegung. Zeitung fur Psychologie 61, 161–265.

Xiao, D., Raiguel, S., Marcar, V., Koenderink, J., Orban, G., 1995. Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. Proc. Natl. Acad. Sci. USA 92, 11303–11306.